AD-A063 010 AMERICAN INSTITUTES FOR RESEARCH WASHINGTON D C F/G 5/10

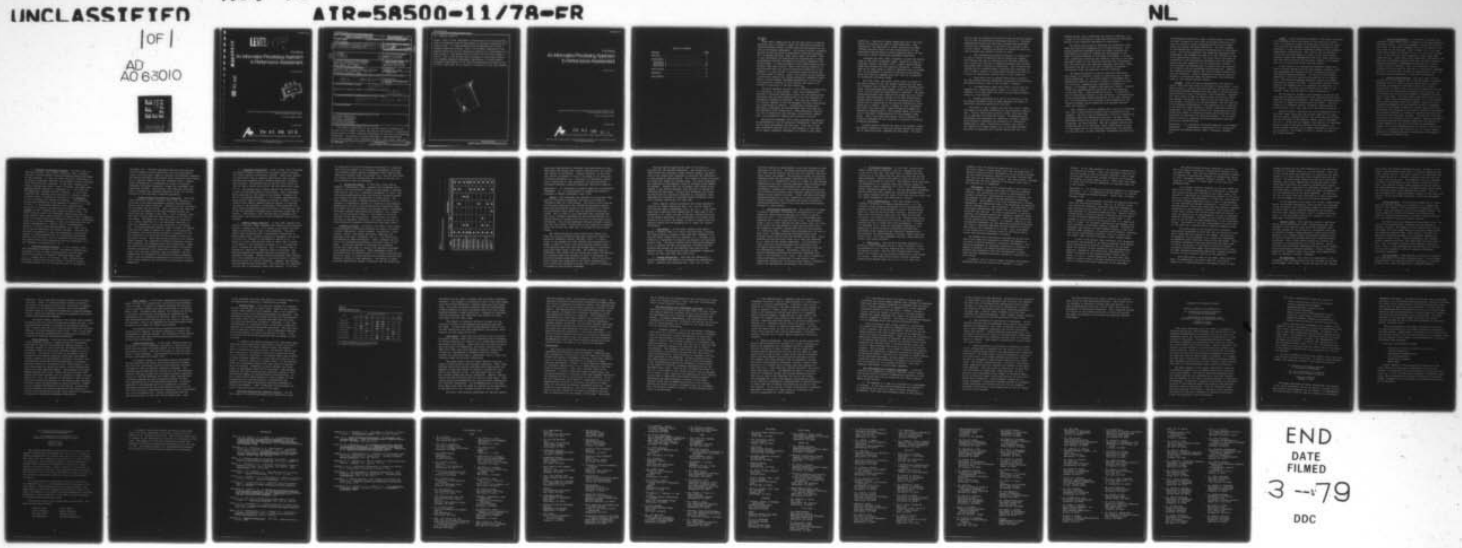
AN INFORMATION PROCESSING APPROACH TO PERFORMANCE ASSESSMENT.(U)

NOV 78 A M ROSE N00014-76-C-0871

UNCLASSIFIED AIR-58500-11/78-FR NL

1 OF 1

AD A063010



END

DATE FILMED

3 --79

DDC

AIR 58500-11/78-FR

LEVEL II

12

SC

AD A063010

DDC FILE COPY

Final Report

An Information Processing Approach to Performance Assessment

Andrew M. Rose

Prepared for the Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research, Arlington, Virginia.

Contract No. N00014-76-C-0871

November 1978

79 01 08 052

AMERICAN INSTITUTES FOR RESEARCH / 1055 Thomas Jefferson Street, NW, Washington, DC 20007

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
Final Report		

4. TITLE (and Subtitle)

6 An Information Processing Approach to Performance Assessment.

5. TYPE OF REPORT & PERIOD COVERED

9 Final Report.
16 Nov [redacted]977 - 30 Nov [redacted]8,

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)

10 A. M. Rose

8. CONTRACT OR GRANT NUMBER(s)

15 N00014-76-C-0871

9. PERFORMING ORGANIZATION NAME AND ADDRESS

American Institutes for Research
1055 Thomas Jefferson St., N.W.
Washington, D.C. 20007

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

61153N; RR042-04-01;
RR042-04; NR 150-391

11. CONTROLLING OFFICE NAME AND ADDRESS

Personnel and Training Research Programs,
Office of Naval Research - Code 458
Arlington, Virginia 22217

12. REPORT DATE

November 30, 1978

13. NUMBER OF PAGES

46

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

14 AIR-58500-11/78-FR

15. SECURITY CLASS. (of this report)

UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

11 30 Nov 78

Approved for public release; distribution unlimited.

12 52p.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

16 RR04204 . 17 RR0420401

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Personnel Technology
Individual Differences
Personnel Assessment
Information Processing
Cognitive Processing

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report summarizes two and one-half years of research concerned with an attempt to apply the theories and methods of laboratory-based studies of human cognitive performance to the area of performance assessment. The primary rationale developed in these experiments (Rose and Fernandes, 1977; Fernandes and Rose, 1978; Allen, Rose, and

DD FORM 1 Jan 73 1473 023 450

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

JUB

20. Abstract (Cont'd.)

Kramer, 1978) is that individuals can potentially be characterized in terms of parameters derived from models of selected information processing tasks. If these parameters can be demonstrated to meet standard test-item criteria, then a test battery comprised of such measures would not only be potentially predictive of performance on a wide variety of real-world tasks but would also be firmly based in theory. Such a test battery would represent a significant advance over standard personnel assessment instruments; it would promote increased understanding of the cognitive operations involved in any criterion task shown to be related to constructs in the test battery.

AIR 58500-11/78-FR

Final Report

An Information Processing Approach to Performance Assessment

Andrew M. Rose

Prepared for the Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research, Arlington, Virginia.

Contract No. N00014-76-C-0871

November 1978

AMERICAN INSTITUTES FOR RESEARCH / 1055 Thomas Jefferson Street, NW, Washington, DC 20007

TABLE OF CONTENTS

Overview

This report summarizes two and one-half years of research concerned with an attempt to apply the theories and methods of laboratory-based studies of human cognitive performance to the area of performance assessment. The primary rationale developed through an ongoing series of experiments (Rose and Fernandes, 1977; Fernandes and Rose, 1978; Allen, Rose, and Kramer, 1978) is that individuals can potentially be characterized in terms of parameters derived from models of selected information processing tasks. If these parameters can be demonstrated to meet standard test-item criteria, then a test battery comprised of such measures would not only be potentially predictive of performance on a wide variety of real-world tasks but would also be firmly based in theory. Such a test battery would represent a significant advance over standard personnel assessment instruments; it would promote increased understanding of the cognitive operations involved in any criterion task shown to be related to constructs in the test battery.

The basic approach involved in this research is exemplified in a precursor, to this project. Rose (1974) employed the strategy of selecting experimental tasks from the psychological literature that had been demonstrated to be valid measures of information processing constructs. Each task was adapted to fit logistic demands of time and equipment and then administered to a large group of subjects. Correlational analyses were conducted to determine the relationships among the tasks and the individual task reliabilities. These procedures resulted in a set of tasks which were reliable, statistically independent, and considered to possess high construct validity.

The Rose and Fernandes (1977) study extended this approach by hypothesizing a set of constructs ("operations") which were used to model performance for each task. Since most of the task parameters could be cast as time measures, it was possible to employ regression techniques to "converge" upon these

operations. Some fairly simple assumptions led to the estimation of durations for some of these operations. More importantly, the generation of these constructs provided a valuable heuristic device for the interpretation of task performance and provided an initial empirical basis for the isolation of basic information processing components.

Fernandes and Rose (1978) attempted to extend the methodology to the realm of memorial tasks. Based on a study by Underwood, Boruch, and Malmi (1977), several memory-related tasks were selected for more detailed evaluation. Although modeling of these tasks in terms of operations was not attempted, it was possible to examine the obtained relationships among the tasks for commonalities that could be interpreted in information processing terms.

The Allen, Rose, and Kramer (1978) study is similar to the above studies in that the general approach was the same: the literature was reviewed in order to select candidate paradigms, these paradigms were adapted to meet logistic limitations, and the tasks were administered to subjects. The major differences between this study and the others are that first, several tasks previously included in the test battery were readministered, primarily to test for alternate-form consistencies and to capitalize on the previous findings for interpretation of results. Second, a number of "new" treatments were built into this study. These treatments extended the theoretical underpinnings of the selected tasks, thus allowing for "stronger" interpretations of the phenomena under study. Third, this study made greater use of analysis of variance techniques for the isolation of potential individual difference paramters.

Although specific details for each of the studies varied, the general approach to fulfilling project objectives consisted of three activities, repeated for each study. The first

activity was an extensive literature review and initial screening of tasks and information-processing constructs. Relevant experimental studies and texts were reviewed and individual contacts with researchers in the field were pursued as part of this review. Candidate paradigms and constructs were further evaluated with several criteria in mind:

1. The information-processing construct or concept had to have a history of empirical and/or theoretical support. The interest here was in constructs that had been developed over a period of time and in research paradigms that had been replicated under a variety of conditions. This criterion was relaxed only in instances where a paradigm was considered to be a "classic" measure of a particular construct but where no evidence of replication could be found in the literature.

2. There had to be an adequate theoretical rationale for the paradigm actually measuring the particular information-processing construct that it was intended to measure.

3. The experimental task itself had to be one that was adaptable to a paper-and-pencil format, to a small digital computer, or to some other form that could be easily administered in a group setting.

4. Enough performance data had to be available so that preliminary estimates could be made regarding the extent of individual variation expected for the task.

The result of the screening activity was a set of tasks that seemed to be prime candidates for more extensive examination. For the second major activity, these tasks were adapted or modified into practical formats. The methodological refinements were evaluated in a series of in-house, informal pilot studies to determine the feasibility of alternative adaptations of tasks, instructions, stimuli, and timing. At the completion of these studies, all of the tasks had been evaluated to determine their logical feasibility and, to a

limited degree, their reliability and construct validity. As a result of this evaluation, tasks were retained and considered worthy of more extensive experimental investigation.

The third activity was to determine the properties of the tasks when they were assembled into a research battery. The primary questions addressed during this phase concerned the replicability of previous findings and the adequacy of the tasks to provide measures of individual differences. In addition, information concerning the construct validity of the tasks and sample norms for the resultant measures were investigated. With the relatively large data base employed, additional issues concerning the ability of the set of measures to separate individuals within the population could be examined.

The approach adopted to validity warrants elaboration. The concept of construct validity is relatively new in experimental psychology. At its current stage of development and mathematical analysis, construct validity is primarily a question of belief, dependent upon researcher's judgments of support or nonsupport stemming from empirical results. Nunnally (1978) has suggested general procedures for the generation of relevant data. These procedures involve: (1) specification of observables relevant to the construct; (2) determination of the relationship between observables of the same construct; and (3) determination of the extent to which measures of the construct produce results predicted from accepted theories about the construct.

Thus, construct validity depends upon a chain of inferences, each link of which relies primarily upon interpretation and judgment. The first link is essentially a series of theoretical hypotheses about the underlying constructs. As such, these hypotheses reflect the authors' particular theoretical biases, vocabulary, and task analyses. The next judgment concerns the interpretation of the individual tasks' group effects as more or less supportive of the underlying operational descriptions. For the most part, we have considered "phenomenon replicability"

as presumptive evidence for these interpretations; confidence has been increased not only from the present results but also from the results of other investigators who have performed empirically-based converging operations. The next judgment is the designation of measures as reflecting one or more operations. For the many measures that adequately represent task performance, the judgment was made as to the relevance of each to the operational construct. The final step in the chain of inferences is the correlational hypothesis that two measures sharing the same operation will be statistically related. If each parameter was hypothesized to measure only one operation, the evidence coud be interpreted straightforwardly. However, the evidence becomes shakier when both parameters measure more than one operation. Without assumptions concerning relative weights or correlations among the operations, the interpretations of the evidence becomes indirect.

Summary. Given the above considerations, the approach implies that each task would be evaluated in three areas. First, where relevant, a primary question would be the replicability of previously-obtained phenomena using the same or similar paradigm. Second would be a more "traditional" test evaluation, concerned with such issues as ease of administration and scoring, equipment demands, efficiency (in terms of time to administer and task length), reliability of task performance, and the character of the response distributions in the population as an indicator of the ability of each measure to uncover individual difference parameters. The third area would be the issues previously mentioned with regard to construct validity and theoretical interpretations of individual and group performance. The following sections review or summarize the three experiments conducted during this research project.

Experiment 1: An Information Processing Approach to Performance Assessment: 1. Experimental Investigation of an Information Processing Performance Battery (Rose & Fernandes, November 1977)

Method. As per the above discussion, three major activities were undertaken during the course of this particular effort. First, an extensive literature review was conducted and evaluative criteria were applied; these procedures resulted in the selection of a set of tasks considered worthy of more extensive experimental investigation. This initial set was reduced to a final set of eight tasks as a result of the second major activity, namely the implementation and pilot study evaluation. The third major activity was the actual conduct of the experiment and the evaluation of the obtained results.

As the result of an interservice agreement between the Office of Naval Research and the Army Research Institute (ARI), this experiment was conducted at ARI's computer-controlled Information Systems Laboratory. The heart of this computer system was a CDC 3300, specifically modified for the experiment. The software developed permitted subjects to proceed individually through the tasks. Instructions, practice blocks stimuli, and feedback were all automated and presented via a CRT display. Individual trial responses were timed (to the nearest 3 msec) and recorded for later transcription.

The 54 subjects (students from Georgetown University who were paid for their participation) were administered the eight tasks on two separate occasions. Each session was approximately two hours in length and scheduled two days apart. Different forms (e.g., different stimuli, different randomizations of stimulus order, etc.) were used in the two testing sessions.

Specific details concerning the actual implementations of the eight tasks, along with the relevant empirical and theoretical support, descriptive data, and other detailed information can be found in the Technical Report (Rose & Fernandes, op. cit.). As one form of summary, each task will be described below in terms of the major expected group phenomena, the actual results obtained, and a summary statement of the logistic evaluation. The reader is urged to consult the Report for further details.

1. Letter Classification. In this task, subjects were required to make same-different judgments to two simultaneously-presented letters. The major phenomena involved are that the time taken to make these judgments differ as a function of the rule given for the decision: "name" matches (e.g., A a) take longer than "physical" matches (e.g., A A), and "rule" matches (e.g., if the rule was "both consonants") take longer than "name" matches. These results were replicated; furtmermore, the magnitude of obtained response times were virtually identical despite differences in implementation between the original studies (Posner & Mitchell, 1967) and the present version. Logistically, this task was excellent--it was easy to implement and score, was efficient in terms of administration time, and produced reliable parameters of individual differences.

2. Lexical Decision Making. The basic task, from the perspective of the subjects, was to decide whether a visually-presented string of letters was an English word or nonword. Following the procedures employed by Meyer (e.g., Meyer, Schvaneveldt, & Ruddy, 1974), each trial consisted of the presentation of two successively-presented letter strings. The two strings bore a particular graphemic or phonemic relationship to each other: they were either physically similar (e.g., COUCH - TOUCH) and/or phonemically simular (e.g., they rhymed). The original phenomena were that graphemic similarity alone inhibited performance and that phonemic as well as graphemic simularity facilitated recognition. While these results were replicated, the magnitude of the effects was small; furthermore, the parameter selected to reflect the "encoding facilitation" effect was only marginally reliable. Logistically, this task was difficult to implement on the computer, requiring carefully-controlled sitmulus and response timing and large amounts of storage capacity for the on-line processing requirements. However, for future implementations, the "simple" form of the task (i.e., presenting one letter string at a time, thereby ignoring the graphemic-phonemic manipulations) was deemed to be a potentially valuable task.

3. Graphemic and Phonemic Analyses. The basic task, as developed by Baron (1973; Baron & McKillop, 1975), required subjects to decide whether visually-presented phrases made sense or were nonsense. In order to study individual differences in the speed of phonemic (acoustic) and graphemic (visual) analysis, three conditions were required. In the first condition two kinds of phrases were used: sense (S) phrases, and those which sounded sensible because of a homophone (e.g., IT'S KNOT SO) but looked like nonsense (called H phrases). In this first condition (SH), subjects were instructed to classify a phrase as making sense or nonsense on the basis of its appearance (so that H phrases were judged as nonsense). The second condition used H phrases and true nonsense (N) phrases (e.g., NEW I CAN'T). In this second condition (HN), subjects were instructed to classify the phrases on the basis of how they sounded, so that H phrases were judged as making sense. The third condition used S and N phrases; subjects were free to choose whatever basis they preferred for making S and N judgments. Since the original study was concerned with individual differences, there are no "group" phenomena to serve as comparisons with the present implementation. However, some of the parameters selected to reflect condition performance (e.g., mean RT for the SN and SH conditions) were highly reliable. This task was difficult to implement in that the development of sufficient numbers of stimulus phrases for all conditions required substantial time and effort, with no assurance that the phrases were equated along other potentially relevant dimensions (e.g., word frequency, pronounceability, etc.).

4. Short-term Memory Scanning. In this "classic" paradigm developed by Sternberg (1967), the general procedure was to present a list of items for memorization that was short enough to be within the subjects' immediate memory span (i.e., a "memory set" of 1-4 digits). Next, subjects were presented a probe digit and were asked to decide as quickly as possible whether or not the probe was one of the items in

the memory set. The basic phenomena were that the functions relating response time to memory set size are approximately linear, and with equal slopes for positive and negative responses. The results obtained in the present study replicated these results in all important aspects (the only exception being a steeper slope, probably due to differences in amount of practice). Individual difference parameters (slopes and intercepts) were moderately reliable. Logistically, this task was relatively easy to implement but required additional effort for the computation of the individual difference parameters.

5. Memory Scanning for Words and Categories. This task, first employed by Juola & Atkinson (1971), was identical in structure to the Short-term Memory Scanning task described above, with one major exception: instead of digits, the memory set and probe item were words. Again, subjects were required to memorize the memory set and then determine whether the probe item was a member of that sets. Juola & Atkinson (1971) also used a second condition that differed from the first in one aspect: the memory set items were names of categories (e.g., COLOR, RELATIVE, etc.) and the probe items were (or were not) exemplars of one of the categories in the memory set. The basic phenomena showed a linear increase in response time with the number of memory set items in both conditions. Furthermore, the functions had equivalent intercepts for the two conditions, but the slope was much steeper for the categorization trials. These results were obtained in the present study; however, the individual difference parameters (slopes and intercepts) had disappointingly low reliabilities. These low reliabilities suggested that subjects might have adopted different strategies on the second day of testing. Logistically, this task was relatively easy to implement; the sole drawback was the generation of sufficient category exemplars (again, with the possible contamination of linguistic factors such as word frequency and difficulty of ascertaining category membership).

6. Linguistic Verification. Clark & Chase (1972) developed and tested a model to account for how people compare information from linguistic and pictoral sources. Their model applied to a particular type of sentence verification task in which subjects were presented with a display containing a sentence and a picture. The sentence was of the form "star (plus) is (is not) above (below) plus (star)" and the picture was either $\overset{*}{+}$ or $\overset{+}{*}$. The subjects had to decide whether the sentence was a true or false description of the picture. Their 4-parameter model accounted for the latencies of subjects' judgments for all of the possible sentences. In the present implementation, group results compared favorably with those of Chase & Clark in terms of both overall response time and pattern of latencies for each of the sentences. Furthermore, model parameter estimates closely paralleled previous findings. Unfortunately, two of the four parameter estimates were unreliable across days of testing. Logistically, this task was easy to implement and administer; the calculations of model parameters were straightforward.

7. Semantic Memory Retrieval. In this task, adapted from Collins & Quillian (1969), subjects were presented with sentences such as "A canary can fly" or "A canary is an animal" and were asked to ascertain the truth of the statements. The basic results were consistent with a hierarchical organization in memory semantically "nested" categorical statements took longer to verify as a function of the degree of nesting (e.g., "A canary is a bird" required less time to verity than "A canary is an animal"), and "property" sentences (e.g., "A canary has skin") followed the same pattern. Thus, if response time is plotted as a function of "required level of hierarchy", the obtained property and superset functions were parallel, with different intercepts. Present results replicated these findings and closely mirrored the Collins & Quillian results in terms of absolute values for function paramters. However, the obtained slope parameters were unreliable. The difficulty with implementing this task is conceptual rather than logistic: it is hard

to argue that an a priori hierarchy constructed for a specific topic would be common to all subjects, furthermore, the stimulus sentences require specific assumptions about factual knowledge that subjects might not possess (e.g., the fact that birds are animals and that animals have skin might not be part of a given subject's store of knowledge).

8. Recognition Memory. In this task, first used by Shepard & Teghtsoonian (1961), subjects were presented with a lengthy list of three-digit numbers and were asked to identify each item as "old" (i.e., previously presented) or "new". The lists were constructed so that the intralist intervals between the original and test presentations of items varied. The basic finding is that a standard retention function (i.e., a decrease in probability correct as a function of time between presentation and test) could be derived for individual items. This result was also obtained in the present study; furthermore, individual subject signal-detection parameters (i.e., proportion of hits and false alarms, d') were shown to be fairly reliable across testing sessions. Logistically, this task was easy to implement, administer, and score; calculation of signal-detection parameters required some additional effort.

Construct Validity. For each of the tasks in this study, it was possible to describe performance in terms of a series of eight "operations", namely encoding, constructing, transforming, storing, retrieving, searching, comparing, and responding (see Rose & Fernandes, op. cit., for definitions and examples of these operations). Table 1 presents an overview of the hypothesized operations involved in each of the tasks. It was also possible to describe each parameter derived from these tasks in terms of the operations presumed to be sampled by the measures. These hypotheses, taken in conjunction with the obtained intra- and intertask correlations, served as inputs to various speculative analyses concerning the construct validity of the operations. The results of these analyses supported the general conclusion that the theoretical

TABLE 1.
Overview of Task Operations in Experiment 1

TASKS	ENCODE	CON-STRUCT	TRANS-FORM	STORE	RETRIEVE	SEARCH	COMPARE	RESPOND
	OPERATIONS							
LETTER CLASSI-FICATION–PHYS.	●						●	●
LETTER CLASSI-FICATION–NAME	●				●		●	●
LEXICAL DE-CISION MAKING	●	●				●		●
GRAPHEMIC/PHONE-MIC ANALYSES	●	●				●		●
STM SCANNING	●						●	●
MEMORY SCAN-NING FOR WORDS	●						●	●
MEMORY SCAN-NING FOR CAT.	●				●		●	●
LINGUISTIC VERIFICATION	●	●	●				●	●
SEMANTIC MEMORY RETRIEVAL	●	●			●			●
RECOGNITION MEMORY	●			●	●		●	●

operations hypothesized to determine task performance do, both empirically and inductively, account satisfactorily for significant aspects of performance. Equally important, it was clear that further refinement of each of the steps in the construct-validation procedures (i.e., definition of operations, assignment of operations to task parameters, and the statistical techniques employed) requires substantial additional research.

Experiment 2: An Information Processing Approach to Performance Assessment: II. An Investigation of Encoding and Retrieval Processes in Memory (Fernandes & Rose, 1978)

Method. This experiment was designed to investigate other types of information processing activities that might be included as part of a test battery. The focus in this case was on structural features of the information processing system, those that describe the nature of the information at a particular processing stage rather than the operations being performed. The six tasks in this experiment were concerned with the nature of memory representation and provided measures of various aspects of encoding and retrieval of previously stored information. This second experiment was more limited in scope than the first study, focusing more on the logistics of administering and scoring the tasks than on reliability and validity issues.

The tasks selected for pilot testing were chosen from among the various recognition-type and recall-type tasks presentend by Underwood, Boruch, and Malmi (1977). In their study, it was assumed that when subjects were presented with a number of words to learn, they would abstract certain kinds of information about each word and perhaps about its relationships with other words in the task. The different types of information about words that get stored were called "attributes", and different tasks were selected in order to determine the interrelationships among memory attributes. Some of the attributes focused upon properties of the stored representation, while others were concerned with how a new chunk of information is integrated into previous knowledge.

The same general approach was used in this study as described for the previous experiment. The initial set of candidate tasks were modified to meet logistic demands, and small-scale pilot studies were conducted to ascertain the feasibility of more extensive administration. As a result of these activities, six tasks were chosen for further study. The actual experiment was conducted on AIR premises, using 22 volunteer staff members as subjects. Subjects were run in groups; each subject participated in two testing sessions, each two hours in length and scheduled two days apart. Stimuli were presented (and timing was controlled by a projector connected to a peripheral timer. Responses were recorded in individual subject booklets, which also contained instructions for each task.

Again, specific details concerning the actual implementations of the tasks, descriptive data, etc., can be found in the Technical Report (Fernandes & Rose, op. cit.). Since these tasks were primarily direct adaptations of tasks used in the Underwood et al. study, and due to the more limited objectives of the present experiment, the presentation of the major group results will be brief; the reader is urged to consult both the Technical Report and the Underwood et al. report for further information.

1. <u>Free Recall</u>. In this task, subjects were shown a series of 20-word lists at a rate of one word every 2 seconds. The lists differed in content: some lists were composed of all concrete nouns, others were all abstract (i.e., presumed to be difficult or impossible to visualize as objects), and some were control lists. The major findings (which replicated Underwood et al.'s results) was that the proportion of words correctly recalled was greater for the concrete lists than for the abstract lists. Reliabilities of basic parameters (proportion correct for each list type) were quite high.

2. <u>Running Recognition</u>. This task was essentially a Recognition Memory task derived from Shepard and Teghtsoonian (1961). Subjects were presented with a long list of items

and were required to judge whether or not they had seen an item previously in the list. The major difference between the present task and the previous one was that the stimulus items were words rather than numbers. The principal finding for both the Underwood study and the present experiment was that recognition functions (i.e., probability correct vs. number of items between presentation and test) could be generated; these functions show a gradual diminution of performance with increasing lags. Furthermore, recognition of word stimuli was more accurate and less affected by increasing test lags when compared to previous research using numbers as stimuli. Three individual-difference parameters-proportion of hits, proportion of false alarms, and overall proportion correct had high test-retest reliability (although performance levels were quite high and some subjects produced no false alarms).

3. Interference Susceptibility. This task was an attempt to measure individual differences in susceptibility to interference by associations established in a series of paired-associate lists. A list consisted of five word-number pairs presented for a single study and test trial. The procedure within a set of lists remained the same across lists; the lists would contain the same words but they would be paired with the numbers in different combinations and would be presented in a different order. Subjects were presented with six sets of such lists. The expectation was that performance would decrease within each set and also decrease across sets. These expectations were partially borne out in both the Underwood et al. study and the present experiment: the proportion of items correctly recalled decreased with successive lists (collapsed across sets); however, a decrease across sets was not consistently obtained. Furthermore, subjects in the present experiment performed the task less accurately than those in the Underwood et al. study. Also, while a measure of overall performance (proportion of items correct) had a respectable reliability, the derived slope measure (intended to reflect the hypothesized interference) was unreliable.

4. Situational Frequency. In this task, subjects were shown a long list of words at a rate of 2 seconds per word. At the end of the list, they were given a response form containing all of the words from the list plus some that had not been presented. Subjects were asked to judge the actual frequency of occurrence of each word. The results, which replicated the Underwood et al. findings, were that subjects tended to overestimate actual frequencies of 0 and 1, but underestimated actual frequencies greater than 1. The two behavioral parameters derived from this task, namely the correlation between actual and judged frequency and the slope of the line relating actual and and judged frequencies, were both highly reliable.

5. List Differentiation. This task focused on the "temporal" attribute of a memory representation, i.e., the ability to order incoming information on the time dimension. Subjects were shown three successive lists of 20 four-letter words at a rate of one word every 2 seconds. Subjects were cued orally and visually when each list ended and the next one began. At the end of the third list, subjects were given a response sheet containing the 60 words and were required to indicate the list in which each word had appeared. In both the Underwood et al. study and the present experiment, the proportion of words correctly classified decreased with successive lists, indicating that subjects' judgments were more accurate for words presented in earlier rather than more recent lists. The single dependent variable, mean proportion correct, was highly reliable.

6. Memory Span. This standard paradigm consisted of a string of letters presented one at a time to subjects; after the presentation of a string, subjects were required to recall the letters in order. The two task variables were string length (6-9 letters) and the acoustic similarity of the letters in a string. High acoustically similar strings contained letters which sounded alike (e.g., B, C, E, G), while the

letters in the low similarity tests did not (e.g., J, R, L, Q). The basic results in both studies were that the proportion of letters recalled correctly decreased as string length increased and that high-acoustically similar strings were recalled less well than the low similarity strings. The latter effect, although reliable was less than expected in the present study than in the Underwood et al. experiment.

Conclusions. Due to the limited scope of this study (in terms of number of subjects and range of cognitive processes involved), the conclusions are limited to considerations of logistic feasibility, replicability of previous findings, and the adequacy of the tasks to provide measures of individual differences. With respect to the first consideration, logistic feasibility, the adaptations made in the procedures and materials used by Underwood et al. were successful. The tasks were easily and quickly administered and scored, and the subjects understood what was required. In terms of replication of previous findings, the obtained results were generally compatible where such comparisons were appropriate. There were some exceptions to this generalization, however. For example, subjects in the Interference Susceptibility task apparently operated differently (in terms of strategies employed and hence pattern of results) than those in the Underwood et al. study. Also, differences in the construction of stimulus lists for the Running Recognition task obviated comparisons of results.

The between-task correlations of parameters derived from each of the tasks were examined to evaluate the adequacy of the tasks to provide measures of individual differences. Such issues as redundancy of variables, sensitivity of measures to subject strategies, and empirically-obtained correlation patterns were examined in the final determination of which tasks and task parameters would be retained for inclusion in a test battery.

In summary, five of the six tasks (excluding Interference Susceptibility) met the criteria for inclusion in a test

battery. All of them appeared to be related to general skill in encoding and storage. Thus, this experiment achieved its desired outcome in that the results indicated a set of tasks and measures which provide reliable estimates of individual differences in general memory skills. These tasks were added to those from the previous experiment as candidate tasks for a test battery.

Experiment 3: "An Information Processing Approach to Performance Assessment: III. An Elaboration and Refinement of an Information Processing Performance Battery" (Allen, Rose, & Kramer, November 1978).

Method. This experiment continued the approach described above for the previous studies. That is, an initial literature review was conducted, candidate tasks were implemented and evaluated in pilot studies, and a final set of eight tasks were selected for large-scale administration. Based on several considerations, including logistic constraints and the desire to test alternate implementations of certain tasks, it was decided to convert each of the selected tasks to a "paper-and-pencil" format. All necessary instructions and response forms were contained in individual subjects' booklets. Although all responses were written on these response sheets, it was still possible (when necessary) to externally pace subjects' responses and control interstimulus intervals, since all instructions and stimuli were carefully recorded on audio cassette tapes.

A total of 68 subjects (paid volunteers from Georgetown University) participated in two testing sessions, each approximately three hours in duration and scheduled one day apart. Again, for specific details concerning task implementations, empirical information, and theoretical considerations, the reader is referred to the Technical Report (Allen, Rose, & Kramer, op. cit.). Before summarizing the individual task results, a brief review of the data analysis activities will serve to reiterate the basic objectives of the experiment.

The general analytical plan consisted of two stages. The first stage was primarily concerned with analyses of the individual tasks. Each task was examined to determine whether the expected (from previous findings) or hypothesized (based on "new" treatments) phenomena actually occurred. This first stage was, in essence, a "forms check" for the particular implementations.

As a general analysis, analysis of variance (ANOVA) was used in this stage. The purpose of the ANOVA was to describe and confirm the previous findings on each task, namely, the pattern of significant and nonsignificant effects of the treatments on overall task performance. In addition, since some of the tasks included repetitions within a day, it was possible to test treatment-by-subject interactions. Significant treatment-by-subject interactions would mean that subjects responded differently to the treatments; therefore, this interaction effect would indicate that further study would be required in order to identify two or more parameters for use in describing subjects' differences. An ANOVA for each task was performed both on the raw data and where appropriate, on the transformed scores. The reason for the data transformation was that some of the tasks had a limited range of possible response scores.

The second stage of the analysis was to estimate individuals' parameters on each task, such as slopes and intercepts, based upon the results of the ANOVAs. The selection of parameters to be estimated (e.g., slopes and intercepts) was dependent upon significant effects from the ANOVAs. After estimating parameters, they were correlated with each other. In theory, the pattern of correlations would show higher correlation coefficients among those parameters which involve the same information processing operations, thereby providing evidence for the construct validity of the operation.

More specifically, three of the tasks, namely Physical Match, Set Membership, and Letter Rotation, were direct adaptations of tasks previously investigated as part of this

research program and have been demonstrated to be potentially valuable as candidate tasks for the test battery. Their use in the current study was essentially to investigate the replicability of previous findings using different materials and formats. Two other tasks, namely Scan and Search and Mental Addition, were direct adaptations of paradigms used by other investigators who were not primarily concerned with issues relating to test construction or individual differences. Thus, for these tasks, a principal concern was again the demonstration of replicability of the major phenomena. In addition, all five of these tasks were evaluated as to their usefulness for the development of individual difference parameters.

Of the remaining three tasks, two of them, namely Letter Recall and Sentence Recognition, while derived from paradigms in the literature, were sufficiently unique in this implementation to merit fairly detailed examination and analysis. The final task, Sentence Recall, was of a slightly different sort: it was developed primarily as a potential source of individual difference parameters.

Physical Match. This task was a partial replication of the Posner and Mitchell (1967) paradigm which was also included in the previous study (Rose and Fernandes, 1977). The main purpose for its inclusion in the present study was to ascertain whether a paper-and-pencil format would produce baseline (i.e., physical match) performance compatible with previous findings from studies which used alternate testing procedures. Such a result would potentially increase the flexibility of administration for a test battery. Results obtained from the present administration were clearly compatible with previously-obtained findings, in terms of absolute magnitude of response times, error rates, and standard deviations of individual subjects' responses.

Set Membership. This task was a replication of the Sternberg (1975) paradigm which was also included in the Rose and Fernandes (1977) study. As was the case for the Physical

Match task, the primary consideration was to determine if the present implementation would result in effects compatible with other testing formats. The basic finding to be replicated was the demonstration of a linearly-increasing function relating time per item to set size. This expected result was obtained. In terms of the parameters of the function (slope and intercept), the results from the three studies were highly similar. For other descriptive measures, the present study and the Rose and Fernandes results were likewise very similar. Thus, it was concluded that a paper-and-pencil version of this task could produce results which were equivalent to previous implementations.

Letter Rotation. This particular adaptation of a paradigm developed by Shepard and co-workers (e.g., Shepard and Metzler, 1971; Snyder, 1972) was previously employed by Rose (1974), using the identical stimuli and response formats. The primary empirical finding of interest was the previously observed, monotonically-increasing function relating time per item and degrees of stimulus rotation required.

The basic result was obtained in the present study: the time per item monotonically increased with the degrees of required rotation. The shapes of these functions are clearly bow-shaped, with the larger degrees of rotation showing faster than expected response times. These results (including the bow-shaped function) mimicked in all important aspects the findings of Rose (1974). In both experiments, the obtained slopes, standard deviations, and ranges were virtually identical. Also, no subject in either study produced a negative slope. Thus, it was concluded that the paradigm had retained its value as a potential source of important individual-difference parameters.

Scan and Search. This task was a variation of Neisser's (1967) procedure for estimating scanning rate. In the present implementation, the task was a direct replication of one used by Rose (1974), with the addition of a degraded stimulus

condition. Thus, there were two major reasons for including this task: first, to confirm previous findings concerning the effect of set size on search rate; and second, to explore the effects of an additional condition (clear and degraded stimuli), both as an indicant of additional processing demands and as a source of individual difference parameters that would reflect these additional demands.

The major results indicated a very strong effect showing processing time per item to increase with target set size. Also, degraded stimuli tended to increase the processing time per item. Results from the clear stimulus conditions were virtually identical to those obtained by Rose (1974). Thus, it was again concluded that this paradigm would be potentially valuable as a source of individual difference parameters.

Mental Addition. This task was an adaptation of Hitch's (1978) paradigm; as such, a primary concern of the present study was to replicate the previously reported results. Of principal interest in this regard was the demonstration of an increasing number of errors with increasing number of carry operations. In addition, this task was included to test an extension of the theory concerning processing requirements. Further evidence for the presence of storage, retrieval, and transformation operations could be obtained by a slightly different casting of the observed data. With respect to evidence concerning the replication issue, Hitch's findings were obtained in every important aspect: the problems with larger addends produced more errors (fewer correct positions; more blanks) and errors increase with an increase in number of carry operations. Thus, for purposes of replication, there is good reason to believe that subjects in the present study performed the task in much the same manner as Hitch's subjects. Regarding the extended interpretation of the results, analyses and results provided very strong evidence that the recasting of the data could potentially provide "good" measures of individual differences in information-processing operations.

Letter Recall. In this task, subjects were required to recall the last five letters from a series that varied between 5 and 10 items. A number of different paradigms and theories lead to consistent predictions about performance in the Letter Recall task. The memory span studies, the studies and theories of proactive interference, and theoretical explanations of displacement (e.g., Atkinson and Shiffrin, 1968) all suggest that performance should deteriorate as the length of the series is increased. Thus it was expected that mean performance would decrease as the number of letters was increased.

The obtained results were as expected from predictions, reflecting the subjects' increasing inability to control the displacement or updating process. The regularity of the results suggested that this task could provide interesting parameters of individual differences.

Sentence Recognition. In this task, subjects were presented with a list of sentences, followed by a second list for which they were asked to decide whether or not they had heard the sentences before. They were also asked to rate the confidence of their judgments. This procedure was a partial replication of Bransford and Franks (1971).

Of primary interest in the present context was whether the data exhibited the effect found by Bransford and Franks, who found that confidence ratings on new-consistent sentences increased in value as sentence complexity increased. The obtained results indicated that ratings for the new-consistent condition did not increase with increasing sentence complexity. Failure to find the Bransford and Franks effect required that a different approach be used to determine whether the data were representative of the abstraction operation. A reinterpretation of the paradigm led to the generation of various hypotheses concerning the relative proportion correct for the four sentence types (consistent-new, consistent-old, inconsistent-new, and inconsistent-old). Results recalculated from the data were entirely in line with the new predictions. Thus,

it was concluded that this task might still provide support for the hypothesized operation of abstraction processes.

Sentence Recall. This task was included for different primary purposes than the other tasks. Based on our literature review, there seemed to be an apparent "gap" in the research concerning recall and clustering processes for sentences. And, although the stimuli were derived partially from the Bransford and Franks (1971) work, the tasks developed here were theoretically and practically different. Therefore, a "new" task was developed as a potential source of individual difference measures of information-processing operations. The basic task was for subjects, after hearing a list of sentences organized around four topics, to recall the sentence by topic. Potential individual difference parameters were evaluated in later analyses.

In summary, the results indicated that, where applicable, the major group effects were replicated in almost every paradigm. For two of the tasks, Sentence Recall and Physical Match, replication was not an issue. The Sentence Recall task was a new task of our design so there was no previous literature of results to replicate. The Physical Match task involved a single treatment condition, so this task could only be compared to previous results at a general level. Of the remaining tasks only the Sentence Recognition task failed to replicate previous findings. Therefore, we were confident that these tasks, as implemented, were "solid" paradigms. They produced phenomena that were consistent with previous findings or were demonstrably capable of straightforward interpretations. As such, these results added additional logical support for the interpretation of task performance as representing the hypothesized information processing operations presumed to underlie performance. Table 2 summarizes the hypotheses concerning these operations for each of the tasks.

Individual measures and construct validity. For the eight tasks, 59 individual parameters were selected to study

TABLE 2.
Operational Analysis of Tasks

Tasks	Information Processing Operations								
	Encoding	Abstraction	Transformation	Recoding	Storage	Retrieval	Search	Comparison	Decision-response
1. Letter recall	minor	minor	minor	minor	**major**	**major**	minor	minor	minor
2. Mental addition	minor	minor	**major**	minor	**major**	**major**	minor	minor	minor
3. Sentence recall	minor	**major**	minor	minor	**major**	**major**	minor	minor	minor
4. Sentence recognition	minor	**major**	minor	minor	**major**	minor	**major**	minor	minor
5. Letter rotation	minor	minor	**major**	minor	minor	minor	minor	**major**	minor
6. Physical match	minor	minor	minor	minor	minor	minor	minor	**major**	minor
7. Set membership	minor	minor	minor	minor	**major**	minor	**major**	minor	minor
8. Scan and search	minor*	minor	minor	minor	**major**	minor	**major**	minor	minor

major = operation is of MAJOR importance in determining task performance.
minor = operation is of MINOR importance in determining task performance.
* For degraded stimuli, encoding is of MAJOR importance

the ability of the tasks to produce good individual difference measures. With respect to reliability, many of the tasks showed good test-retest consistency, although there were some exceptions. The tasks involving processing time per item measures produced moderate to strong reliability while the other tasks produced moderate to weak reliability. Perhaps this indicates that the processing time tasks were less subject to coding strategies than the other tasks.

With regard to construct validity, evidence for the construct validity of the hypothesized operations was good but mixed. In many cases where we hypothesized tasks to use the same operation(s) we found consistently significant correlations; in other cases we did not.

Conclusions. Based upon the various criteria outlined previously, certain tasks from this experiment were considered as strong candidates for inclusion in an information processing performance test battery. These tasks were the Physical Match, Scan and Search, Set Membership, and Letter Rotation tasks. These four tasks replicated published findings, showed good reliability, and appeared to possess construct validity. Further, the tasks were easy to administer, easy to score, and alternate forms can be easily constructed.

The Letter Recall task and Mental Addition task comprised a second group of more marginal candidates. These tasks produced significant effects consistent with expectations and appeared to have some degree of construct validity but reliability was moderate to low. The tasks were easy to score and alternate forms were easily constructed. However, administration was more difficult than for the four tasks above. Both tasks required good mechanisms for timing stimulus presentation and intervals consistently. The recommendations in the Technical Report were that the tasks be included in any application of the battery but that attempts be made to improve them.

The final, most marginal candidates for the test battery

were the Sentence Recall and Sentence Recognition tasks. The Sentence Recognition task produced significant treatment effects but failed to replicate the desired finding. The task produced very low reliability and alternate forms were difficult to construct. The Sentence Recall task produced significant treatment effects but showed low to moderate reliability and was also moderately difficult to construct. The difficulty in construction for these tasks stemmed from a difficulty in controlling extraneous variables such as sentence complexity, sentence length, vocabulary, and any other variable that might have influenced subject strategies from sentence to sentence. The tasks also were moderately difficult to administer, requiring moderate consistency in the timing of events. Construct validity appeared to be good, but scoring, especially for the Sentence Recall task, was complex and time consuming. Recommendations were that these tasks be replaced with other tasks designed to converge on the abstraction operation.

Implications

This research project encompasses several usually independently-investigated research domains. Depending upon one's perspective, this project could be viewed as a straightforward test battery development exercise, a data base compilation, or an empirical investigation of theories of human cognitive performance. Clearly, it would have been impossible to thoroughly explore any or all of these domains in the time frame of this project. Thus, there are many issues still unresolved. In order to delineate some of these issues, the following discussion is organized around the problems confronting two different potential utilizations of this research. First, we will discuss pure "test battery" issues--what would be the considerations involved in the direct utilization of these tasks in a particular performance assessment situation. These issues put aside theoretical concerns of construct validity; that is, we will assume for the sake of discussion that the user is satisfied as to the content of the tests. The second

set of issues will be concerned with the utilization of these tasks as a research battery. Here, the major considerations center around construct validity.

The test battery as an assessment instrument. In any particular utilization of one or more tasks, the user is faced with several practical problems, having to do primarily with the implementation and logistics of administering the task battery. Equipment requirements, time considerations, and procedures for instructing, collecting and scoring are critical. Each of these problems, as it relates to the tasks studied, will be discussed individually.

1. Equipment demands. As one of the primary considerations during our initial screening and evaluation of candidate tasks, we arbitrarily limited the selection to tasks which could be administered to groups of subjects and would thus require at worst a small digital computer, or preferably only paper and pencils. At one time, the goal was to make the task battery completely portable for potential use in "exotic" environments. The equipment constraints have their manifestations mainly in the control the user has over the timing of stimulus presentation, timing accuracy of responses, and automatic recording of the responses. Previous research using the tasks studied in this project did not (with minor exceptions) systematically explore the necessity of such controls or the impact of different equipment implementations on the phenomena under study. While we feel confident that most (if not all) of the tasks in the current battery could be implemented in several forms, empirical confirmations are still lacking. We were able to demonstrate equivalences between computer-implemented and paper-and-pencil versions of some tasks (Physical Match and Set Membership directly in the current research and several others by inference--that is, by comparisons with previous research); however, further demonstrations remain to be performed.

2. Time considerations. Depending upon the specific assessment situation, time for administration of a task battery might be a critical consideration. A parallel issue is the question of "efficiency" - how much time does a task require in order to generate a useful datum? We did not explore these issues, and arbitrarily limited task length on a judgmental basis, since we deemed it inappropriate at this stage of the research to consider such issues as test length vs. score quality, numerical efficiency scores, or alternate experimental designs (e.g., "Bayesean testing" where the amount of data collected is a function of subjects' performance). The resolution of these issues, as well as other time-related considerations (total testing time, warm-up requirements, intertask spacing, fatigue, etc.), would require substantial additional research.

3. Testing procedures. This group of issues-instructions, ease of data collection and scoring--in a sense transcends any particular implementation, and perhaps should be addressed during a discussion of test validity. We have found that instructions are particularly critical for the types of tasks studied here, since we want to make fairly precise inferences about just what subjects are doing during individual trials. For example, the Letter Rotation task is supposed to measure the rate at which subjects manipulate (rotate) a mental representation; if a subject were to adopt some other approach to this task (e.g., a non-rotational feature analysis), the results would lose their value. Our approach has been to "structure" subjects' strategies through instructions whenever possible. Thus, we are open to the objection that we are not testing true processes but rather task-specific requirements; hence, we are seriously compromising interpretations of generalizability to other tasks and ultimately the predictive validity of the task battery. In our opinion, careful investigation of subjects' approaches (strategies) to these tasks is a neglected and critical requirement for future research.

At a more mundane level, the issues of ease of data collection and scoring depend somewhat upon equipment limitations. Some tasks would become infeasible if equipment were not available to record responses or if generation of scores required inordinate amounts of time and effort. However, the most crucial issue with regard to scoring is a validity question: whether the scores generated for a task truly measure what they are supposed to measure. The choice of appropriate measures will be addressed during the validity discussion below.

There are several other issues related to task battery utilization that were partially addressed during the course of this research, but due to the scope of the project and the unique nature of the tasks studied could not be explored as fully as we would have preferred. These include test reliability, population characteristics of measures, and practice effects. It is obvious that further research is required to more firmly establish test reliabilities, to enlarge the data base, and to more clearly determine the effects of practice on task performance. This statement is not meant to downplay the absolute necessity and importance of empirical research on reliability and practice effects, especially for the domain of tasks studied. As was mentioned previously, we were unable to find in the literature (with some exceptions) any mention of reliability and no references at all to practice effects. Thus, this research requirement is particularly important.

<u>The task battery as a research instrument</u>. In the discussion of our approach to construct validity, several issues were raised which we consider fundamentally important not only for this project but also as objectives for future research. These issues include:

1. The adequacy of a task as a reflection of a phenomenon of interest. Do the task results demonstrate the presence of an unambiguous and nontrivial aspect of human information processing? This issue has manifested itself in this project

in the criteria for task selection. We restricted our selection to tasks where the group phenomena were striking; the argument was that even if the theoretical explanations might be questionable, at least the data would be unambiguous, reliable, and potentially explainable given other tasks administered at the same time. Clearly, these judgments would have been different for other researchers. Thus, another research requirement is the exploration and incorporation of other tasks.

2. Task parameters as reflections of phenomena. Do the measures derived from the paradigms truly capture the critical aspects of performance? This issue is probably much deeper and of more fundamental importance than we have envisioned during this project. Issues regarding scale properties, distributions, and so forth, were not systematically investigated, nor were implications of different scoring rules for interpretation of results. These remain as potentially critical areas for future study.

3. Task parameters as reflections of hypothetical operations. Perhaps more fundamental are the issues of the definitions and hypothesized nature of the operations used to describe performance. We intentionally have not attempted to develop a full-scale model of the human information-processing system, nor have we attempted to exhaustively sample all such operations. The current research project is at best a preliminary study, especially when one considers the genesis of the operations employed. Over the course of the project, we have modified definitions of operations, added new ones, and have generally been unsystematic with respect to designing experiments (or selecting tasks) which would provide critical information about operations. Such systematic studies are, of course, necessary to eventually provide evidence on the hypothisized constructs, and would be critical for any research concerned with building a test battery such as the one conceptualized in this project.

One final implication for future research is so obvious that it has not been mentioned previously. What stands out is the necessity to externally validate the task battery. It must be demonstrated that this (or any other) test predicts performance in operational situations. This validation against a criterion is the ultimate objective at which this project was aimed, and remains as the single most pressing research requirement.

ABSTRACTS OF TECHNICAL REPORTS

Technical Reports AIR-58500-TR-1-2-3

An Information Processing Approach
to Performance Assessment:

I. Experimental Investigation of an
Information Processing Performance Battery

Andrew M. Rose
Kathleen Fernandes

This report describes the first study in a program of research dealing with the development and validation of a comprehensive standardized test battery that can be used as an assessment device for the evaluation of performance in a wide variety of situations. The standardized battery is being designed to possess high reliability and predictive validity for a wide variety of criterion tasks. Equally important, the battery is being designed to include tests that possess construct validity: there will be a firm theoretical and empirical base for inferring the information possessing structures and functions that the tests purport to measure. It is expected that such a battery will permit improved personnel management decisions to be made for a wider variety of Navy-relevant jobs than is currently possible using existing techniques.

The major purpose of the present experimental study was to determine properties of the tasks selected for inclusion in the test battery. The primary questions addressed during this phase concerned the replicability of previous findings and the adequacy of the tests to provide measures of individual differences. In addition, information concerning the construct validity of the tasks and population norms for the resultant measures could be investigated. With the relatively large data base employed (54 subjects), additional data concerning the ability of the set of measures to separate individuals within the population could be examined.

The tests investigated included the following:

Letter classification (Posner and Mitchell)
Lexical decision making (Meyer)
Graphemic/phonemic analyses (Baron)
STM scanning (Sternberg)
Memory scanning for words/categories (Juola)
Linguistic verification (Clark and Chase)
Recognition memory (Shepard and Teghtsoonian)
Semantic memory retrieval (Collins and Quillian)

Several questions were addressed in this phase of the research. First, replicability of previous experimental work with similar paradigms was investigated. In general, the results were quite compatible with previous findings for all eight tasks. The second area addressed concerned the established of the reliability, validity, and independence of the tasks being studied. In general the reliabilities for most measures was quite high ($r \geq .50$). The measures were also analyzed to determine practice effects and the character of the response distributions in the population for each of the measures.

In order to address validity-type issues, inter- and intra-task correlations were calculated. In general, these analyses support ehc construct validity of the tasks and measures.

An Information Processing Approach to Performance Assessment:

II. An Investigation of Encoding and Retrieval Processes in Memory

Kathleen Fernandes
Andrew M. Rose

This paper describes the tasks, methodology, and results of the second experiment carried out during a research program dealing with the development and validation of a comprehensive test battery which could be used as a research or performance

assessment instrument. The focus in this case was on structural features of the information processing system, those that describe the nature of the information at a particular processing stage rather than the operations being performed. The six tasks in this experiment were concerned with the nature of memory representation and provided measures of various aspects of encoding and retrieving of previously stored information. The tasks selected for testing were chosen from among the various recognition-type and recall-type tasks presented by Underwood, Boruch, and Malmi (1977).

The major purpose of the present experimental study was to determine properties of the tasks selected for inclusion in the test battery. Specifically, the issues of the replicability of previous findings, the logistic feasibility of our adapta tions, and the adequacy of the tests to provide measures of individual differences were addressed.

The tasks investigated included:

Free recall (control, concrete, and abstract)
Running recognition
Interference susceptibility
List differentiation
Situational frequency
Memory span

Several aspects of the results were evaluated for the final selection of measures to incorporate into the test battery. Consideration of the replicability of previous findings, Day 1 - Day 2 reliabilities, practice, intratask correlations, and the patterns of intertask correlations led to the summary judgments for the inclusion of a relatively small number of variables.

An Information Processing Approach
to Performance Assessment:

III. An Elaboration and Refinement of an
Information Processing Performance Battery

Ted W. Allen
Andrew M. Rose
Leslie Kramer

This report describes the third study in a program of res research regarding the development and validation of a comprehensive standardized test battery that can be used as an assessment device for the evaluation of performance in a wide variety of situations. The standardized battery is being designed to possess high reliability and predictive validity for a wide variety of criterion tasks. Equally important, the battery is being designed to include tests that possess construct validity: there will be a firm theoretical and empirical base for inferring the information processing structures that the tests purport to measure.

The major purpose of the present study was to determine the properties of a set of tasks selected for inclusion in the test battery. Three questions were of primary interest: the replicability of previous findings with alternate forms of the tasks, the adequacy of the tasks to provide measures of individual differences, and the adequacy of the tasks to provide measures of information processing operations.

Sixty-eight subjects were tested twice on each task. The tasks investigated included:

Physical Match	Letter Recall
Letter Rotation	Mental Addition
Scan and Search	Sentence Recall
Set Membership	Sentence Recognition

In general, the results showed the forms of tasks used here to be quite compatible with previous findings for all tasks but one, Sentence Recognition. Even for this task, there is support in the experimental literature for the obtained findings. The support for individual difference measures and measures of information-processing operations varied from task to task. The summary presents our recommendations regarding the inclusion of various tasks and measures in the battery.

REFERENCES

Allen, T. W., Rose, A. M., & Kramer, L. An information processing approach to performance assessment: III. An elaboration and refinement of an information processing performance battery. Washington, D.C.: American Institutes for Research, 1978.

Atkinson, R. C. & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (eds.), The psychology of learning and motivation: Advances in research and theory, 2, New York: Academic Press, 1968, pp. 89-195.

Baron, J. Phonemic stage not necessary for reading. Quarterly Journal of Experimental Psychology, 1973, 25, 241-246.

Baron, J. & McKillop, B. J. Individual differences in speed of phonemic analysis, visual analysis, and reading. Acta Psychologica, 1975, 39, 91-96.

Bransford, J. D. & Franks, J. J. The abstraction of linguistic ideas. Cognitive Psychology, 1971, 2, pp. 331-350.

Clark, H. H. & Chase, W. G. On the process of comparing sentences against pictures. Cognitive Psychology, 1972, 3, 472-517.

Collins, A. M. & Quillian, M. R. Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 240-247.

Fernandes, K. & Rose, A. M. An information processing approach to performance assessment: II. An investigation of encoding and retrieval processes in memory. Washington, D.C.: American Institutes for Research, 1978.

Hitch, G. J. The role of short-term working memory in mental arithmetic. Cognitive Psychology, 1978, 10, pp. 302-323.

Juola, J. F. & Atkinson, R. C. Memory scanning for words versus categories. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 522-527.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. Functions of graphemic and phonemic codes in visual word-recognition. Memory and Cognition, 1974, 2, 309-321.

Neisser, U. Cognitive Psychology. New York: Appleton-Century-Crofts, 1967.

Posner, M. I. & Mitchell, R. F. Chronometric analysis of classification. Psychological Review, 1967, 74, pp. 392-409.

Rose, A. M. Human information processing: An assessment and research battery. Human Performance Center Technical Report No. 46. Ann Arbor, Michigan, 1974.

Rose, A. M. & Fernandes, K. An information processing approach to performance assessment: I. Experimental investigation of an information processing performance battery. Washington, D.C.: American Institutes for Research, 1977.

Shepard, R. N. & Teghtsoonian, M. Retention of information under conditions approaching a steady state. Journal of Experimental Psychology, 1961, 62, 302-309.

Shepard, R. N. & Metzler, J. Mental rotation of three-dimensional objects. Science, 1971, 171, pp. 701-703.

Snyder, C. R. R. Individual differences in imagery and thought. Ph.D. Dissertation, University of Oregon, 1972.

Sternberg, S. Two operations in character recognition: Some evidence from reaction-time measurements. Perception and Psychophysics, 1967, 2, pp. 45-53.

Sternberg, S. Memory scanning: New findings and current controversies. Quarterly Journal of Experimental Psychology, 1975, 27 pp. 1-32.

Underwood, B. J., Boruch, R. F., & Malmi, R. A. The composition of episodic memory. Evanston, Illinois: Northwestern University, 1977.

Distribution List

NAVY

1 Dr. Ed Aiken
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Jack R. Borsting
Provost & Academic Dean
U.S. Naval Postgraduate School
Monterey, CA 93940

1 Dr. Robert Breaux
Code N-71
NAVTRAEQUIPCEN
Orlando, FL 32813

1 Mr. Maurice Callahan
Pers 23a
Bureau of Naval Personnel
Washington, DC 20370

1 Chief of Naval Education and
Training Support - (01A)
Pensacola, FL 32509

1 Dr. Richard Elster
Department of Administrative
Sciences
Naval Postgraduate School
Monterey, CA 93940

1 Dr. Pat Federico
Navy Personnel R&D Center
San Diego, CA 92152

1 CDR John Ferguson, MSC, USN
Naval Medical R&D Command
Code 44
National Naval Medical Center
Bethesda, MD 20014

1 Dr. Richard Gibson
Bureau of Medecine and Surgery
Code 513
Navy Department
Washington, DC 20372

1 Capt. D.M. Gragg, MC, USN
Head, Section on Medical Education
Uniformed Services University of
the Health Sciences
6917 Arlington Road
Bethesda, MD 20014

1 Mr. George N. Graine
Naval Sea Systems Command
Sea 047C112
Washington, DC 20362

1 Dr. Steve Harris
Code L522
MANRL
Pensacola, FL 32508

1 LCDR Charles W. Hutchins
Naval Air Systems Command
444 Jefferson Plaza #1
1411 Jefferson Davis Highway
Arlington, VA 20360

1 CDR Robert S. Kennedy
Naval Aerospace Medical and
Research Lab
Box 29407
New Orleans, TN 38054

1 Dr. Norman J. Kerr
Chief of Naval Technical
Training
Naval Air Station Memphis (75)
Millington, TN 38054

1 Dr. Leonard Kroeker
Navy Personnel R&D Center
San Diego, CA 92152

1 Chairman, Leadership & Law
Department
Division of Professional
Development
U.S. Naval Academy
Annapolis, MD 21402

1 Dr. William L. Maloy
Principal Civilian Advisor for
Education and Training
Naval Training Command
Code 00A
Pensacola, FL 32508

1 Capt. Richard L. Martin
USS Francis Marion (LPA-Z49)
FPO New York, NY 09501

1 Dr. James McBride
Code 301
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. William Montague
LRDC
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

1 Commanding Officer
U.S. Naval Amphibious School
Coronado, CA 92155

1 Commanding Officer
Naval Health Research Center
Attn: Library
San Diego, CA 92152

1 Naval Medical R&D Command
Code 44
National Naval Medical Center
Bethesda, MD 20014

1 Capt. Paul Nelson, USN
Chief, Medical Service Corps
Code 7
Bureau of Medicine & Surgery
U.S. Department of the Navy
Washington, DC 20372

1 Library
Navy Personnel R&D Center
San Diego, CA 92152

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 Office of Civilian Personnel
Code 26
Department of the Navy
Washington, DC 20390

1 John Olsen
Chief of Naval Education & Training Support
Pensacola, FL 32509

1 Psychologist
ONR Branch Office
495 Summer Street
Boston, MA 02210

1 Psychologist
ONR Branch Office
536 S. Clark Street
Chicago, IL 60605

1 Office of Naval Research
Code 200
Arlington, VA 22217

1 Office of Naval Research
Code 441
800 N. Quincy Street
Arlington, VA 22217

5 Personnel & Training Research Programs
Code 458
Office of Naval Research
Arlington, VA 22217

1 Psychologist
Office of Naval Research Branch
223 Old Marylebone Road
London, NW, 15th England

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Scientific Director
Office of Naval Research
Scientific Liaison Group/Tokyo
American Embassy
APO San Francisco, CA 96503

1 Head, Research, Development and Studies (OP102X)
Office of the Chief of Naval Operations
Washington, DC 20370

1 Scientific Advisor to the Chief of Naval Personnel (Pers-Or)
Naval Bureau of Personnel
Room 4410, Arlington Annex
Washington, DC 20370

1 Dr. Richard A. Pollak
Academic Computing Center
U.S. Naval Academy
Annapolis, MD 21402

1 Mr. Arnold Rubenstein
Naval Personnel Support Technology
Naval Material Command (08T244)
Room 1044, Crystal Plaza #5
2221 Jefferson Davis Highway
Arlington, VA 20360

1 Dr. Worth Scanland
Chief of Naval Education and Training
Code N-5
NAS, Pensacola, FL 32508

1 A.A. Scjoholm
Technical Support
Code 201
Navy Personnel R&D Center
San Diego, CA 92152

1 Mr. Robert Smith
Office of Chief of Naval Operations
OP-987E
Washington, DC 20350

1 Dr. Alfred F. Smode
Training Analysis & Evaluation Group
(TAEG)
Department of the Navy
Orlando, FL 32813

1 CDR Charles J. Theisen, Jr. MSC, USN
Head Human Factors Engineering Division
Naval Air Development Center
Warminster, PA 18974

1 W. Gary Thomson
Naval Ocean Systems Center
Code 7132
San Diego, CA 92152

1 Dr. H.M. West III
Deputy ADCNO for Civilian Planning and Programming
Room 2625, Arlington Annex
Washington, DC 20370

1 Dr. Martin F. Wiskoff
Navy Personnel R&D Center
San Diego, CA 92152

ARMY

1 ARI Field Unit-USAREUR
Attn: Library
c/o ODCSPER
HQ USAREUR & 7th Army
APO New York 09403

1 Technical Director
U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 HQ USAREUE & 7th Army
ODCSOPS
USAAREUE Director of GED
APO New York 09403

1 Dr. Ralph Canter
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Milton S. Katz
Individual Training & Skill
Evaluation Technical Area
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Harold F. O'Neil, Jr.
Attn: PERI-OK
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. James L. Raney
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Director, Training Development
U.S. Army Administration Center
Attn: Dr. Sherrill
Ft. Benjamin Harrison, IN 46218

1 Dr. Joseph Ward
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

AIR FORCE

1 Air Force Human Resources Lab
AFHRL/PED
Brooks AFB, TX 78235

1 Air University Library
AUL/LSE 76/443
Maxwell AFB, AL 36112

1 CDR Mercer
CNET Liaison Officer
AFHRL/Flying Training Division
Williams AFB, AZ 85224

1 Personnel Analysis Division
HQ USAF/DPXXA
Washington, DC 20330

1 Research Branch
AFMPC/DPMYP
Randolph AFB, TX 78148

1 Dr. Marty Rockway AFHRL/TT
Lowry AFB
Colorado 80230

1 Jack A. Thorpe, Capt., USAF
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

1 Brian K. Waters, LCOL, USAF
Air University
Maxwell AFB
Montgomery, AL 36112

MARINES

1 Director, Office of Manpower Utilization
HQ Marine Corps (MPU)
BCB, Building 2009
Quantico, VA 22134

1 MCDEC
Quantico Marine Corps Base
Quantico, VA 22134

1 Dr. A.L. Slafkosky
Scientific Advisor
Code RD-1
HQ U.S. Marine Corps
Washington, DC 20380

COAST GUARD

1 Mr. Joseph J. Cowan, Chief
Psychological Research (G-P-1/62)
U.S. Coast Guard HQ
Washington, DC 20590

OTHER DOD

1 Dr. Stephen Andriole
Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

12 Defense Documentation Center
Cameron Station, Building 5
Alexandria, VA 22314
Attn: TC

1 Dr. Dexter Fletcher
Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

1 Military Assistant for Training and Personnel Technology
Office of the Under Secretary of Defense for Research and Engineering
Room 3D129, The Pentagon
Washington, DC 20301

1 Major Wayne Sellman, USAF
Office of the Assistant Secretary of Defense (MRA&L)
Room 3B930, The Pentagon
Washington,DC 20301

1 Mr. Fredrick W. Suffa
MPP (A&R)
Room 2B269, The Pentagon
Washington, DC 20301

CIVIL GOVT.

1 Dr. Susan Chipman
Basic Skills Program
National Institute of Education
1200 19th Street N.W.
Washington, DC 20208

1 Dr. Lorraine D. Eyde
Personnel R&D Center
U.S. Civil Service Commission
1900 E Street N.W.
Washington, DC 20415

1 Dr. William Gorham, Director
Personnel R&D Center
U.S. Civil Service Commission
1900 E Street N.W.
Washington, DC 20415

1 Dr. Joseph I. Lipson
Division of Science Education
Room W-638
National Science Foundation
Washington, DC 20550

1 Dr. John Mays
National Institute of Education
1200 19th Street N.W.
Washington, DC 20208

1 National Institute of Education
1200 19th Street N.W.
Washington, DC 20208

1 Dr. Andrew R. Molnar
Science Education Development and Research
National Science Foundation
Washington, DC 20550

1 Dr. H. Wallace Sinaiko
Program Director
Manpower Research and Advisory Services
Smithsonian Institution
801 North Pitt Street
Alexandria, VA 22314

1 Dr. Thomas G. Sticht
Basic Skills Program
National Institute of Education
1200 19th Street N.W.
Washington, DC 20208

1 Robert W. Stump
Eeucation and Work Group
National Institute of Education
1200 19th Street N.W.
Washington, DC 20208

1 Dr. Vern W. Urry
Personnel R&D Center
U.S. Civil Service Commission
1900 E Street N.W.
Washington, DC 20415

1 C.S. Winiewicz
U.S. Civil Service Commission
Regional Psychologist
230 S. Dearborn Street
Chicago, IL 60604

1 Dr. Joseph L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 20550

NON GOVT.

1 Prof. Earl A. Alluisi
Department of Psychology
Code 287
Old Dominion University
Norfolk, VA 23508

1 1 Psychological Research Unit
Department of Defense (Army Office)
Campbell Park Offices
Canberra ACT 2600, Australia

1 Dr. Gerald V. Barrett
Department of Psychology
University of Akron
Akron, OH 44325

1 Dr. Nicholas A. Bond
Department of Psychology
Sacramento State College
600 Jay Street
Sacramento, CA 95819

1 Dr. David G. Bowers
Institute for Social Research
University of Michigan
Ann Arbor, MI 48106

1 Dr. C. Victor Bunderson
WICAT Inc.
University Plaza, Suite 10
1160 S. State Street
Orem, UT 84057

1 Dr. John B. Carroll
Psychometric Lab
University of North Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Charles Myers Library
Livingstone House
Livingstone Road
Stratford
London E15 2LJ England

1 Dr. William Chase
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 Dr. Kenneth E. Clark
College of Arts & Sciences
University of Rochester
River Campus Station
Rochester, NY 14627

1 Dr. Norman Cliff
Department of Psychology
University of Southern California
University Park
Los Angeles, CA 90007

1 Dr. Meredith Crawford
Department of Engineering
Administration
George Washington University
Suite 805
2101 L Street N.W.
Washington, DC 20037

1 Dr. Ruth Day
Center for Advanced Study in
Behavioral Sciences
202 Junipero Serra Boulevard
Stanford, CA 94305

1 Dr. Marvin D. Dunnette
N492 Elliott Hall
Department of Psychology
University of Minnesota
Minneapolis, MN 55455

1 Major I.N. Evonic
Canadian Forces Personnel
Applied Research
1107 Avenue Road
Toronto, Ontario, Canada

1 Dr. Richard L. Ferguson
The American College Testing
Program
P. O. Box 168
Iowa City, IA 52240

1 Dr. Victor Fields
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Edwin A. Fleishman
Advanced Research Resources
Organization
8555 Sixteenth Street
Silver Spring, MD 20910

1 Dr. John R. Frederiksen
Bolt, Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 Dr. Vernon S. Gerlach
College of Education
146 Payne Building B
Arizona State University
Tempe, AZ 85281

1 Dr. Robert Glaser
LRDC
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

1 Dr. James G. Greeno
LRDC
University of Pittsburgh
3939)'Hara Street
Pittsburgh, PA 15213

1 Dr. Ron Hambleton
School of Education
University of Massachusetts
Amherst, MA 01002

2 Dr. Barbara Hayes-Roth
The Rand Corporation
1700 Main Street
Santa Monica, CA 90406

1 Dr. James R. Hoffman
Department of Psychology
University of Delaware
Newark, DE 19711

1 Library
HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921

1 Dr. Earl Hunt
Department of Psychology
University of Washington
Seattle, WA 98105

1 Mr. Gary Irving
Data Sciences Division
Technology Services Corporation
2811 Wilshire Boulevard
Santa Monica, CA 90403

1 Dr. Lawrence B. Johnson
Lawrence Johnson & Assoc., Inc.
Suite 502
2001 S Street N.W.
Washington, DC 20009

1 Dr. Roger A. Kaufman
203 Dodd Hall
Florida State University
Tallahassee, FL 32306

1 Dr. Steven W. Keele
Department of Psychology
University of Oregon
Eugene, OR 97403

1 LCOL C.R.J. LaFleur
Personnel Applied Research
National Defense Headquarters
101 Colonel By Drive
Ottawa, Canada K1A0K2

1 Dr. Alan Lesgold
Learning R&D Center
University of Pittsburgh
Pittsburgh, PA 15260

1 Dr. Frederick M. Lord
Educational Testing Service
Princeton, NJ 08540

1 Dr. Robert R. Mackie
Human Factors Research, Inc.
6780 Cortona Drive
Santa Barbara Research Pk.
Goleta, CA 93017

1 Richard T. Mowday
College of Business Administration
University of Oregon
Eugene, OR 97403

1 Dr. Allen Munro
University of Southern California
Behavioral Technology Labs
3717 South Hope Street
Los Angeles, CA 90007

1 Dr. Donald A. Norman
Department of Psychology C-009
University of California, San Diego
La Jolla, CA 92093

1 Dr. Melvin R. Novick
Iowa Testing Programs
University of Iowa
Iowa City, IA 52242

1 Dr. Jesse Orlansky
Institute for Defense Analysis
400 Army Navy Drive
Arlington, VA 22202

1 Dr. James A. Paulson
Portland State University
P. O. Box 751
Portland, OR 97207

1 Mr. A.J. Pesch, President
Eclectech Associates, Inc.
P. O. Box 178
N. Stonington, CT 06359

1 Mr. Luigi Petrullo
2431 N. Edgewood Street
Arlington, VA 22207

1 Dr. Steven M. Pine
4950 Douglas Avenue
Golden Valley, MN 55416

1 Dr. Frank Pratzner
Center for Vocational Education
Ohio State University
1960 Kenny Road
Columbus, OH 43210

1 Dr. Diane M. Ramsey-Klee
R-K Research and System Division
3947 Ridgemont Drive
Malibu, CA 90265

1 Min. Ret. M. Rauch
P II 4
Bundesministerium Der
Verteigdigung
Postfach 161
53 Bonn 1, Germany

1 Dr. Peter B. Read
Social Science Research Council
605 Third Avenue
New York, NY 10016

1 Dr. Mark D. Reckase
Educational Psychology Department
University of Missouri-Columbia
12 Hill Hall
Columbia, MO 65201

1 Dr. Leonard L. Rosenbaum, Chairman
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Ernst Z. Rothkopf
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

1 Prof. Fumiko Samejima
Department of Psychology
University of Tennessee
Knoxville, TN 37916

1 Dr. Benjamin Schneider
Department of Psychology
University of Maryland
College Park, MD 20742

1 Dr. Walter Schneider
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Dr. Richard Snow
School of Education
Stanford University
Stanford, CA 94305

1 Dr. Robert Sternberg
Department of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520

1 Mr. D.J. Sullivan
c/o Canyon Research Group, Inc.
741 Lakefield Road
Westlake Village, CA 91361

1 Dr. Patrick Suppes
Institute for Mathematical
Studies in the Social Sciences
Stanford University
Stanford, CA 94305

1 Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

1 Dr. Kikumi Tatsuoka
Computer Based Education Research
Laboratory
252 Engineering Research
Laboratory
University of Illinois
Urbana, IL 61801

1 Dr. Perry Thorndyke
The Rand Corporation
1700 Main Street
Santa Monica, CA 90406

1 Dr. Benton J. Underwood
Department of Psychology
Northwestern University
Evanston, IL 60201

1 Dr. Thomas Wallsten
Psychometric Laboratory
Davie Hall 013A
University of North Carolina
Chapel Hill, NC 27514

1 Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455

1 Dr. Susan E. Whitely
Psychology Department
University of Kansas
Lawrence, KS 66044